知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum
E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于专利发明人人名消歧的研发团队识别研究

◎ 张静 [1,2,3] 张志强 [4] 赵亚娟 [1]

[1] 中国科学院文献情报中心 北京 100190
[2] 中国科学院大学 北京 100049
[3] 中国科学院档案馆 北京 100190
[4] 中国科学院成都文献情报中心 成都 610041

摘要：[目的 / 意义] 技术研发的核心是人才。研发团队是各领域技术发展的重点关注对象，也是机构研发实力的重要体现。[方法 / 过程] 以德温特创新索引（DII）专利文献为分析对象，明确发明人人名消歧规则，利用发明人共现聚类确定主要研发团队，然后以 3D 打印的数字光处理相关专利来进行人名消歧后研发团队识别的实证分析。[结果 / 结论] 证明专利发明人人名消歧有利于发明人专利数量的准确分析。

关键词：专利 发明人 研发团队识别 人名消歧
分类号：G353.1 G306

引用格式: 张静, 张志强, 赵亚娟. 基于专利发明人人名消歧的研发团队识别研究 [J/OL]. 知识管理论坛, 2016, 1(3): 217-225[引用日期]. http://www.kmf.ac.cn/paperView?id=42.

技术研发的核心是人才，信息环境下信息量的爆炸式增长使得技术研发更加离不开研究团队的通力协作，在人才引进等具体政策制定上除了关注首席专家，更应关注在研发团队中起到核心作用的关键人才。研发团队识别作为专利分析的重要内容之一，有利于甄别核心团队成员，发现非首席的关键人才，能为政策制定和关键研发人员识别提供更好的支持。但是研发人员姓名具有很强的歧义性，存在同名多指及同人不同写法的歧义问题，因此研发团队识别研究最首要的问题就是进行人名消歧，此时人名消歧的核心目标为保障准确率。

1 人名消歧研究进展

人名消歧主要是对姓名表述相同或相近的两个姓名是否指向同一人作出判断。A. Bagga 等 [1] 于 1998 年就开始把跨文本人名消歧作为一种人名共指问题进行探索。2007 年、2009 年和 2010 年 WePS 评测研讨会进行了针对网络人名消歧的评测。在国内，CIPS-SIGHAN-2012 会议 [2] 对中文人名识别与消歧的研究也越来越多。

基于网页等资源进行人名相关的实体特征抽取、聚类，以进行人名消歧的相关研究较多，同时社会网络、阈值或概率确定原则等也都是人名消歧研究中探索使用的方法。如 G. Mann 等 [3] 在 2003 年通过定制模板来提取网页个人传记特征来构造特征向量的方法对人名进行“消歧”。M. B. Fleischman 等 [4] 在 2004 年抽取名字特征、网页特征、重叠特征、语义特征等，使用最大熵模型来计算两个名字指向同一实体的概

作者简介：张静 (ORCID：0000-0002-5827-7139)，馆员，博士研究生，E-mail：zhangjing@mail.las.ac.cn；张志强，教授，博士生导师；赵亚娟，副研究员，硕士生导师，博士。

收稿日期：2016-03-14 发表日期：2016-06-25 本文责任编辑：王铮



率。B. Malin[5] 于 2005 年提出基于社会网络来进行人名消歧。K. Balog 等 [6] 于 2007 年通过训练好的语言模型计算网页中人名指向某个实体的概率，再确定阈值以实现人名消歧。Y. Chen 等 [7] 在 2007 年通过抽取基于名词短语的特征和命名实体的特征，再使用层次凝聚聚类方法进行聚类。S. Ono 等 [8] 在 2008 年基于命名实体共指、关键词以及主题信息的混合特征来对文档进行聚类。L. Romano 等 [9] 于 2009 年提出 XMedia 系统采用质量阈值聚类算法。章顺瑞等 [10] 于 2010年采用层次聚类算法对中文人名进行消歧。陈晨等 [11] 在 2011 年利用不同社会网络边权值和不同图划分准则对人名消歧效果的影响进行了中文人名消歧的研究。

随着人名消歧研究的不断深入，为提高准确性，针对特定数据源的人名消歧、多种方法结合的分步式研究开始增多。2012 年，杨欣欣等 [12] 利用网络资源用搜索引擎四类查询规则扩展特征文档，利用二层聚类算法 [13] 来进行人名消歧。2013 年李广一等 [14] 根据特征类型来设置权值，进行多次聚类。2014 年 S. Christian 等 [15] 利用数据库文献间的引用构建社会网络图来实现特定数据源的人名消歧。2015 年，阳怡林等 [16] 通过上下文特征、实体特征、社会关系特征，利用 3 种不同的聚类算法得到不同的聚类划分，再最终集成来提高人名消歧的准确性。D. H. Han 等 [17] 采用极限学习机提出了针对每一个姓名及姓名集合的两种聚类算法来进行人名消歧。M. Song 等 [18] 针对 PubMed 数据库构建了专门的训练集，并提出新的出版特征集合以提高准确性。

整体来看，当前研究的主要对象以网络资源或论文著者为主，具体方法上以通过改进算法获取更多人名相关特征，或采用多次 / 多层聚类的方法来进行比对判断为主。这些方法均存在一定程度的人名消歧误差，且这部分误差为算法直接判定得出的结果，分析人员并不确定误差可能涉及的人名范围，因此存在一定的“黑箱”问题。

当前针对专利文献的具体特征进行发明人人名消歧的相关研究较少。而专利发明人的著录方式在不同数据库中有所不同，基本都同时涉及中国人名及外国人名的消歧问题。另外，作为政策支撑的专利发明人人名消歧工作需要在确保准确的核心目标下提高效率。因此基于专利文献的人名消歧需要在明确专利数据库发明人姓名结构特征的基础上进行具体探索，以提升准确性，并减少“黑箱子”问题带来的误差不确定性。

2 专利发明人人名消歧

德温特创新索引（DII）是经过人工智力加工后的专利数据，具有可以批量获取、自然语言检索及不同来源专利数据统一再分类的优势，是专利分析的常用数据之一。本文将针对该数据库，结合汤森创新（TI）专利数据库中发明人机构、国家等特征信息来进行专利发明人人名消歧规则研究。

2.1 人名消歧流程

本文主要通过发明人姓名结构特征来进行姓名相似度比较，然后利用专利文献中可获取的发明人特征信息进行判断，以实现人名消歧，见图 1。

2.2 专利发明人姓名结构特征及影响

不同国家来源发明人姓名的结构特征对人名歧义的影响有所不同。通过实际数据查看，可以发现发明人姓名结构特征主要分为两类：其一是类西方姓名结构；其二是类中国姓名结构。这两种姓名结构特征见表 1。两种姓名的结构特征决定了类西方姓名中出现不同姓名表述同指概率更高，而类中国姓名中出现同样姓名表述却不同指的概率更高。

2.3 专利发明人特征信息

在 DII 与 TI 数据库中涉及到的专利发明人特征信息包括姓名缩写、姓名全称、地址（其中包括发明人国家信息）、所属专利人藏号、所属机构、合作人员等信息（见表 2）。这些信息在数据库中的完备程度有所不同，整体来说：① TI 中的姓名信息完备程度要高于 DII

知识管理论坛，2016(3):217–225
DOI：10.13266/j.issn.2095–5472.2016.026

数据库；② TI 姓名全称字段的取值却有部分专利与姓名缩写取值相同，属于不完备状态；③地址信息中的国家信息完备程度高于城市等信息；④专利人藏号及合作人员信息均较为完备。

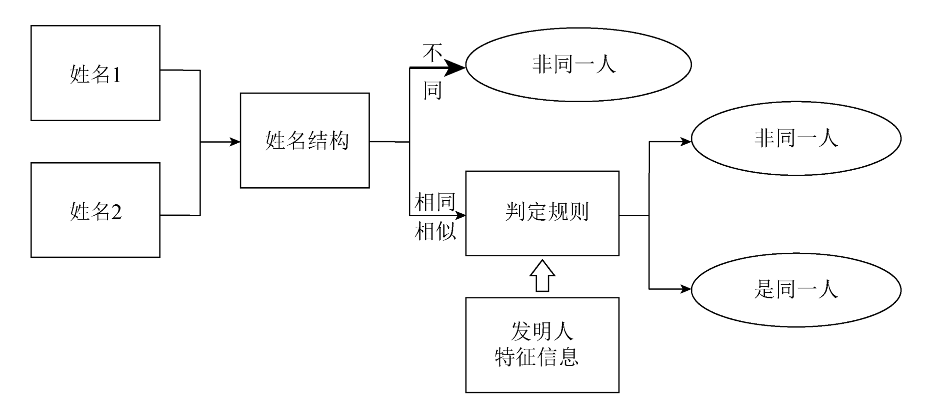


图 1 专利发明人人名消歧流程

表 1 专利发明人姓名结构特征

姓名结构		DII 数据库姓名处理	对人名歧义的影响
类西方姓名	名中间名姓	名：保留全称 中间名：保留全称 / 首字母缩写 / 不体现 姓：首字母缩写 / 不体现	不同姓名表述同指的概率更高
类中国姓名	姓名	姓：保留全称 名：首字母缩写	同样姓名表述却不同指的概率更高

表 2 专利发明人特征信息

发明人特征信息	来源数据库 (来源字段)	完备程度	取值示例
姓名缩写	DII(AU)	2	DRIESSEN M
	TI(发明人 -DWPI)	3	DRIESSEN M
姓名全称	TI(发明人 - 原始)	2	Driessen Marcus Matheus
地址	TI(发明人 - 带有地址)	2	Driessen Marcus Matheus,Maasbracht,NL
所属专利入藏号	DII(GA)	3	2010-L63571
	TI(DWPI 入藏号)	3	2010L63571
所属机构	DII(AE)	2	DSM IP ASSETS BV
	TI(专利权人 - 原始)	2	DSM IP Assets B.V.
合作人员	DII(AU)	2	EL-SIBLANI A；SHKOLNIK A
	TI(发明人 - 原始)	3	SHKOLNIK, Alexandr | EL-SIBLANI, Ali

注：完备程度表示在数据库中对所有专利文献该字段有值数据涵盖范围，3 代表全部专利该字段均有值且数据结构较为规范；2 代表少部分专利该字段为空或取值相对不完整。

2.4 人名消歧规则

人名消歧首先要找出那些可能存在疑问的姓名表述方式，这就需要根据专利发明人姓名相似程度来进行判断，具体的判断标准见表 3。需要指出的是，此处的判断不考虑姓名表述中出现的圆点、连词符等符号信息。

基于以上专利发明人姓名结构特征 (见表 1) 及可获取的专利发明人特征信息 (见表 2)，通过实际数据验证，可以按优先级构建出如下类西方姓名及类中国姓名的人名消歧规则。



表 3 专利发明人姓名相似程度判定标准

姓名结构	相似程度	判定标准	示例	
			姓名缩写	姓名全称
类西方姓名	相同	表述完全一致	ALLANIC A L	Allanic Andre Luc
			ALLANIC A L	Allanic Andre Luc
	相似	①名表述一致；②若有中间名，则首字母缩写一样；③若有姓，则首字母缩写一样	CHAWLA C P	CHAWLA C P
			CHAWLA P C	Chawla Chander P
			CHAWLA P	CHAWLA Prakash
			CHAWLA	Chander
	不同但相近	①名表述仅有一个字母不同；②中间名与姓至少有一项；③若有中间名，则首字母缩写一样；④若有姓，则首字母缩写一样	GOIHAIT H	Goihait Hanan
			GOTHAIT H	GOTHAIT HANAN
	不同	非以上 3 种情况		
类中国姓名	相同	表述完全一致	Luo L	Luo Lei
			Luo L	Luo Lei
	相似	姓表述相同；名表述第一个字母相同，后续表述缺少	Li D	——
			Li Dichen	
	不同	名表述首字母不同	Luo L	Luo Lei
			Luo X	Luo Xiao

2.4.1 类西方姓名消歧规则

根据类西方姓名的结构特征，可以明确对此类姓名消歧的重点在于将同一人的多种姓名表述归一为一种表述。因此对类西方姓名的消歧以专利发明人姓名缩写为入口开始，一方面可以尽可能排除非同一人的姓名表述，另一方面也可以将尽可能多的姓名表述纳入进一步判断范畴。具体规则描述如表 4 所示：

表 4 类西方姓名消歧规则

<table>
<tr><th>序号</th><th>姓名缩写</th><th>入藏号</th><th>姓名全称</th><th>国家</th><th>所属机构</th><th>共同合作人</th><th>结论</th></tr>
<tr><td>1</td><td>不同</td><td></td><td></td><td></td><td></td><td></td><td>非同一人</td></tr>
<tr><td>2</td><td rowspan="3">相同</td><td>相同</td><td></td><td></td><td></td><td></td><td>同一人</td></tr>
<tr><td>3</td><td rowspan="2">不同</td><td></td><td>不同</td><td>不同</td><td></td><td>非同一人</td></tr>
<tr><td>4</td><td></td><td></td><td>相同</td><td>有重复</td><td>同一人概率大</td></tr>
<tr><td>5</td><td rowspan="5">相似</td><td>相同</td><td>相同 / 相似</td><td></td><td></td><td></td><td>同一人</td></tr>
<tr><td>6</td><td rowspan="4">不同</td><td rowspan="4">相似</td><td rowspan="3">相同</td><td rowspan="2">相同</td><td>有重复</td><td>同一人</td></tr>
<tr><td>7</td><td>无重复</td><td>非同一人概率大</td></tr>
<tr><td>8</td><td>不同</td><td>无重复</td><td>非同一人概率大</td></tr>
<tr><td>9</td><td>不同</td><td></td><td></td><td>非同一人</td></tr>
<tr><td>10</td><td rowspan="2">不同但相近</td><td>相同</td><td></td><td></td><td></td><td></td><td>非同一人</td></tr>
<tr><td>11</td><td>不同</td><td></td><td></td><td>相同</td><td></td><td>同一人</td></tr>
</table>

知识管理论坛，2016(3):217-225
DOI：10.13266/j.issn.2095-5472.2016.026

2.4.2 类中国姓名消歧规则

根据类中国姓名的结构特征，可以明确对此类姓名消歧的重点在于将不同人同样表述的姓名区分开来。同样选择从姓名缩写为入口开始，以尽可能区分出非同一人的情况。具体规则描述如表 5 所示：

表 5 类中国姓名消歧规则

<table>
<tr><th>序号</th><th>姓名缩写</th><th>入藏号</th><th>姓名全称</th><th>国家</th><th>所属机构</th><th>共同合作人</th><th>结论</th></tr>
<tr><td>12</td><td>不同</td><td></td><td></td><td></td><td></td><td></td><td>非同一人</td></tr>
<tr><td>13</td><td></td><td></td><td>不同</td><td></td><td></td><td></td><td>非同一人</td></tr>
<tr><td>14</td><td></td><td>相同</td><td></td><td></td><td></td><td></td><td>同一人概率大</td></tr>
<tr><td>15</td><td>相同</td><td></td><td></td><td></td><td>相同</td><td>有重复</td><td>同一人</td></tr>
<tr><td>16</td><td></td><td>不同</td><td></td><td></td><td>相同</td><td>无重复</td><td>非同一人概率大</td></tr>
<tr><td>17</td><td></td><td></td><td></td><td></td><td>不同</td><td></td><td>非同一人概率大</td></tr>
<tr><td>18</td><td></td><td></td><td rowspan="2">相同</td><td></td><td>相同</td><td></td><td>同一人</td></tr>
<tr><td>19</td><td>相似</td><td></td><td></td><td></td><td>有重复</td><td>同一人</td></tr>
<tr><td>20</td><td></td><td></td><td>不全 / 不同</td><td></td><td></td><td></td><td>非同一人概率大</td></tr>
</table>

值得注意的是，在对于以上人名进行消歧过程中，结论仅为概率性判定，而非确定性结果的规则，需要给出相关具体条目，进行扩展查询，辅以人工判断来给出最终结论。在完成人名消歧的基础上，可以根据数据情况，按共同拥有专利数量或比例情况来确定不同数据集的主要研发团队判定标准，从而通过专利发明人共现聚类来实现研发团队识别。

3 基于人名消歧的数字光处理研发团队识别实证研究

本文以 3D 打印的数字光处理（Digital Light Process，DLP）技术相关专利为例来进行人名消歧后研发团队识别的具体实证。

3.1 人名消歧数量统计结果对比

经过检索及专家判读后，从 DII 数据库中共获取 DLP 技术相关专利 274 项、810 件。同一批专利经过的温特入藏号及发明人姓名表述去重后，DII 原始数据中共涉及 640 个专利发明人姓名表述，TI 原始数据中共涉及 652 个专利发明人姓名表述，按照 2.4 小节所述规则进行人名消歧，按照 TI 数据中的姓名简称进行统计，发现 DLP 技术的 120 名发明人存在同一人多种姓名表述，共有 90 种姓名表述为多人重名情况，最终确定共有 602 名发明人参与研发。

人名消歧前后主要发明人（参与研发专利数量大于 3 项）及其专利数量分布见表 6。可以看出，通过人名消歧，主要发明人 HULL CHARLES W 的专利数量从 5 项变为 6 项，KRITCHMAN Eliahu M. 的专利数量从 4 项变为 5 项（以上见表 6 中阴影部分），使得主要发明人数量排序及数量统计更为准确。

3.2 人名消歧后研发团队识别研究

在人名消歧的基础上，首先利用 Bibexcel 生成发明人共现矩阵，生成可供可视化的节点数据，然后利用 Pajek 工具得到图 2 所示的发明人聚类网络。DLP 技术领域中的 602 名发明人中共有 63 名发明人参与聚类。根据数据情况，本文定义研发团队中至少需要包括 3 名发明人。



知识管理论坛
2016 年 第 3 期 (总第 3 期)

表 6 DLP 技术相关专利主要发明人统计比对 (专利数量 >3 项)

人名消歧后	所属机构	专利数量（项）	人名消歧前	专利数量（项）
HULL CHARLES W	3D SYSTEMS INC	6	ELSIBLANI A	6
El-Siblani Ali	Global Filtration Systems, Inc. ENVISIONTEC GMBH	6	HULL C W	5
Napadensky Eduardo	STRATASYS INC	5	NAPADENSKY E	5
PARTANEN JOUNI P	3D SYSTEMS INC	5	PARTANEN J P	4
KRITCHMAN Eliahu M.	STRATASYS INC	5	GRELIN J	4
Grelin Jerome	HUNTSMAN	4	LI D	4
LI Di-chen	UNIV XIAN SCI & TECHNOLOGY XIAN JIAOTONG UNIV	4	KRITCHMAN E M	4
Larsen Niels Holm	HUNTSMAN	4	LARSEN N H	4
Hangaard Ole	HUNTSMAN	4	SHKOLNIK A	4
FUJISAWA KAZUTOSHI	SEIKO EPSON CORP	4	FUJISAWA K	4
Shkolnik Alexandr	Global Filtration Systems, Inc. ENVISIONTEC GMBH	4	HANGAARD O	4

从图 2 中可以清楚看到 DLP 技术领域共有来自 6 个机构的 7 个研发团队。表 7 展现了这些研发团队的情况。可以发现，来自 HUNTSMAN 公司的两个研发团队在 DLP 技术领域并无联系人员，因此被明显区分为两个团队；来自 3D SYSTEMS INC 的研发团队共由 11 人组成，这 11 人又可以大致区分为两个团队（在表 7 中用 A、B 表示），两个团队以 HULL CHARLES W 和 PARTANEN JOUNI P 为纽带，在图 2 中呈现为一个大的团簇。

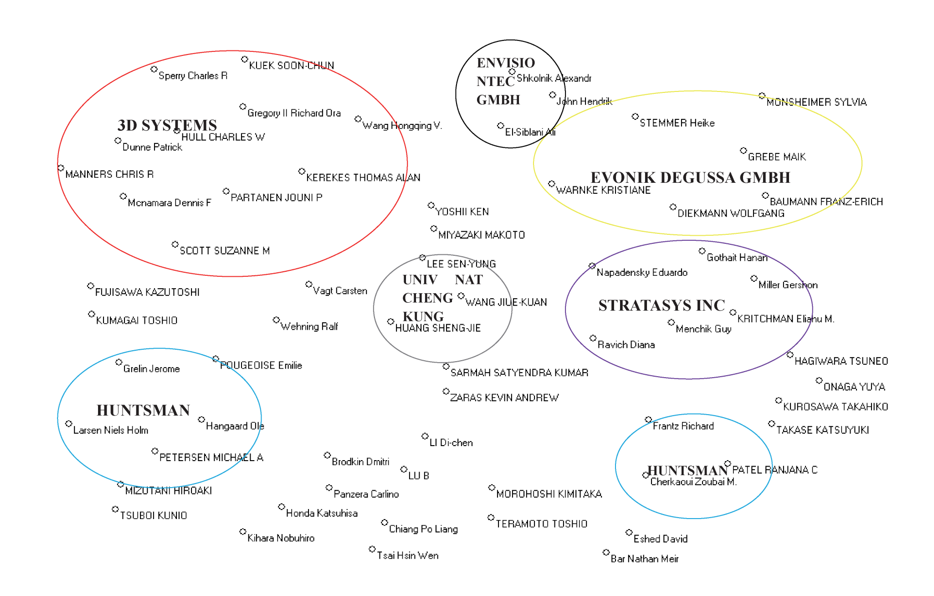


图 2 DLP 技术相关专利发明人共现聚类

知识管理论坛，2016(3):217–225
DOI：10.13266/j.issn.2095–5472.2016.026

表 7 DLP 技术研发团队情况（单位：项）

序号	研发团队 [专利数量]	团队专利数量	所属机构
1	HULL CHARLES W[6] PARTANEN JOUNI P[5] A)Gregory II Richard Ora[3] A)KEREKES THOMAS ALAN[3] A)Wang Hongqing V[3] A)KUEK SOON-CHUN[2] B)MANNERS CHRIS R[2] B)Mcnamara Dennis F[2] B)SCOTT SUZANNE M[2] B)Sperry Charles R[2] B)Dunne Patrick[2]	19	3D SYSTEMS INC
2	Napadensky Eduardo[5] KRITCHMAN Eliahu M.[5] Gothait Hanan[3] Menchik Guy[2] Miller Gershon[2] Ravich Diana[2]	14	STRATASYS INC
3	El-Siblani Ali[6] Shkolnik Alexandr[4] John Hendrik[2]	6	ENVISIONTEC GMBH
4	Grelin Jerome[4] Larsen Niels Holm[4] Hangaard Ole[4] PETERSEN MICHAEL A[2] POUGEOISE Emilie[2]	4	HUNTSMAN
5	BAUMANN FRANZ-ERICH[2] DIEKMANN WOLFGANG[2] GREBE MAIK[2] MONSHEIMER SYLVIA[2] STEMMER Heike[2] WARNKE KRISTIANE[2]	2	EVONIK DEGUSSA GMBH
6	HUANG SHENG-JIE[2] LEE SEN-YUNG[2] WANG JIUE-KUAN[2]	2	UNIV NAT CHENG KUNG
7	Cherkaoui Zoubai M[2] Frantz Richard[2] PATEL RANJANA C[2]	2	HUNTSMAN

3.3 实证研究小结

由于本文人名消歧规则是结合特定数据库的数据结构所提出来的，不具有普适性，因此并未进行人名消歧规则性能测评。

但通过 3.1 部分人名消歧前后主要发明人拥有专利数量对比可以发现，主要发明人的排序有所变化。即通过本文提出的人名消歧，使得主要发明人数量排序及数量统计更为准确，有利于发明人专利数量的准确分析，能够减少由于发明人人名是否同指的不确定性而带来的研发团队识别误差，亦有助于更准确地进行专利研发团队识别。

4 结论

人名消歧结果的准确性将影响到专利分析结果的准确性，从而影响依此为参考而进行的竞争对手识别及相关人才政策决策，因此人名消歧是专利分析不断深入过程中需要解决的重要问题之一。

本文认为，专利研发团队识别过程中的人名消歧应以确保准确性为前提。因此，本文提出的人名消歧规则借鉴了特征向量相似度判定的思路，但在实际操作过程中，具有与其他方法不同的两方面特征：一是基于特定专利数据库数据结构特征来提炼规则，更具有针对性；二是对于无法在逻辑上直接给出确定性结论的规则所涉及条目辅以人工判断来尽量确保准确性，从而避免其他方法直接判定而带来部分不确定性的“黑箱”问题。

本文的人名消歧规则通过实证研究证明是有利于发明人专利数量的准确分析的，但需要指出的是，本文所提出的规则是基于特定专利文献数据的，得出的规则本身在实际应用范围上具有局限性，但针对特定数据而言更具准确性。在今后的研究中，需要进一步探索完善人名消歧方法，扩展人名消歧规则，货站其适用的数据范围，从而更好地进行研发团队识别。

参考文献：

[1] BAGGA A, BALDWIN B. Entity-based cross-document conferencing using the vector space model[C]//COLING’98: Proceedings of the 17th international conference on computational linguistics. New York: ACM Press, 1998: 79-85.

[2] 中国学术会议在线 . 第二届 CIPS-SIGHAN 中文处理国际会议 [EB/OL]. [2014-04-10]. http://www.meeting.edu.cn/meeting/meetingAction-29689!detail.action.

[3] MANN G, YAROWSKY D. Unsupervised personal name disambiguation[C]//CONLL’ 03: Proceedings of the 7th conference on natural language learning at HLT-NAACL 2003. Edmonton: Association for Computational Linguistics, 2003: 33-40.

[4] FLEISCHMAN M B, HOVY E. Multi-document person name resolution[EB/OL]. [2014-03-14]. http://acl.ldc.upenn.edu/W/W04/W04-0701.pdf.

[5] MALIN B. Unsupervised name disambiguation via social network similarity[C]//Proceedings of 2005 SIAM international conference on data mining. Newport Beach: Siam Workshop on Link Analysis, 2005: 93-102.

[6] BALOG K, AZZOPARDI L, RIJKE M D. UVA: Language modeling techniques for web people search[C]//Proceedings of the 4th international workshop on semantic evaluations. Prague: International Workshop on Semantic Evaluations, 2007: 468-471.

[7] CHEN Y, MARTIN J. Towards robust unsupervised personal name disambiguation[EB/OL]. [2014-03-14]. http://acl.ldc.upenn.edu/D/D07/D07-1020.pdf?origin=publication_detail.

[8] ONO S, SATO I, YOSHIDA M, et al. Person name disambiguation in web pages using social network, compound words and latent topics[C]//Proceedings of the 12th pacific-asiaconference on advances in knowledge discovery and data mining. Berlin:Pacific-asiaconference on advances in knowledge discovery and data mining, 2008: 260-271.

[9] ROMANO L, BUZA K, GIULIANO C. XMedia: Web people search by clustering with machinelylearned similaritymeasures[EB/OL]. [2014-03-14]. https://www.researchgate.net/publication/228569058_XMedia_Web_People_Search_by_Clustering_with_Machinely_Learned_Similarity_Measures.

[10] 章顺瑞 , 游宏梁 . 基于层次聚类算法的中文人名消歧 [J]. 现代图书情报技术 , 2010(11): 64-68.

[11] 陈晨 , 王厚峰 . 基于社会网络的跨文本同名消歧 [J]. 中文信息学报 , 2011(5): 76-82.

[12] 杨欣欣 , 李培峰 , 朱巧明 . 基于查询扩展的人名消歧 [J]. 计算机应用 , 2012, 32(9): 2488-2490,2507.

[13] 杨欣欣 , 李培峰 , 朱巧明 . 基于网页文本依存特征的人

知识管理论坛，2016(3):217−225
DOI：10.13266/j.issn.2095−5472.2016.026

名消歧 [J]. 计算机工程 , 2012(19): 133-136.
[14] 李广一 , 王厚峰 . 基于多步聚类的汉语命名实体识别和歧义消解 [J]. 中文信息学报 , 2013, 27(5): 29-34.
[15] CHRISTIAN S, AMIN M, ALEXANDER M P, et al. Exploiting citation networks for large-scale author name disambiguation [J]. EPJ data science, 2014, 3(11): 1-12.
[16] 阳怡林 , 周杰 , 李弼程 . 基于聚类集成的人名消歧算法 [J/OL]. 计算机应用研究 , 2015: 33. [2016-05-30]. http://www.cnki.net/kcms/detail/51.1196.TP.20151028.1121.120.html.
[17] HAN D H, LIU S Q, HU Y C, et al. ELM-based name disambiguation in bibliography [EB/OL].[2016-04-13].http://link.springer.com/article/10.1007%2Fs11280-013-0226-4.
[18] SONG M, KIM E H, KIM H J. Exploring author name disambiguationon PubMed-scale[J]. Journal of informetrics, 2015(4): 924-941.

作者贡献说明：
张静：提出论文写作思路，准备数据，撰写论文；
张志强：完善论文思路，提出修改建议；
赵亚娟：完善论文思路，提出修改建议。

Identification of R&D Teams Based on the Disambiguation of Patent Inventors' Names

Zhang Jing[1,2,3] Zhang Zhiqiang[4] Zhao Yajuan[1]
[1]National Science Library, Chinese Academy of Sciences, Beijing 100190
[2]University of Chinese Academy of Sciences, Beijing 100049
[3]Archives of Chinese Academy of Sciences, Beijing 100190
[4]Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041

Abstract: [Purpose/significance] Talent is the main factor in technology research. R&D teamsare the focus of technology development in various fieldsandimportant manifestation of the competence of an institution. **[Method/process]** Based on Derwent Innovation Index (DII) patent documents, the rules for the disambiguation of patent inventors' names weredefined, and the key R&D teams were identified by inventors clustering. Then, an empirical study was carried out on patents related to Digital Light Process (DLP) of 3D printing. **[Result/conclusion]** It is shown that the disambiguation of patent inventors' names is helpful to accurately analyze the number of inventors' patents.

Keywords: patents   inventors   identification of R&D teams   disambiguation of names